

The State of HDF5 – 2022-3

SC 22, Dallas, TX
November 16, 2022



Dana Robinson
The HDF Group

Overview

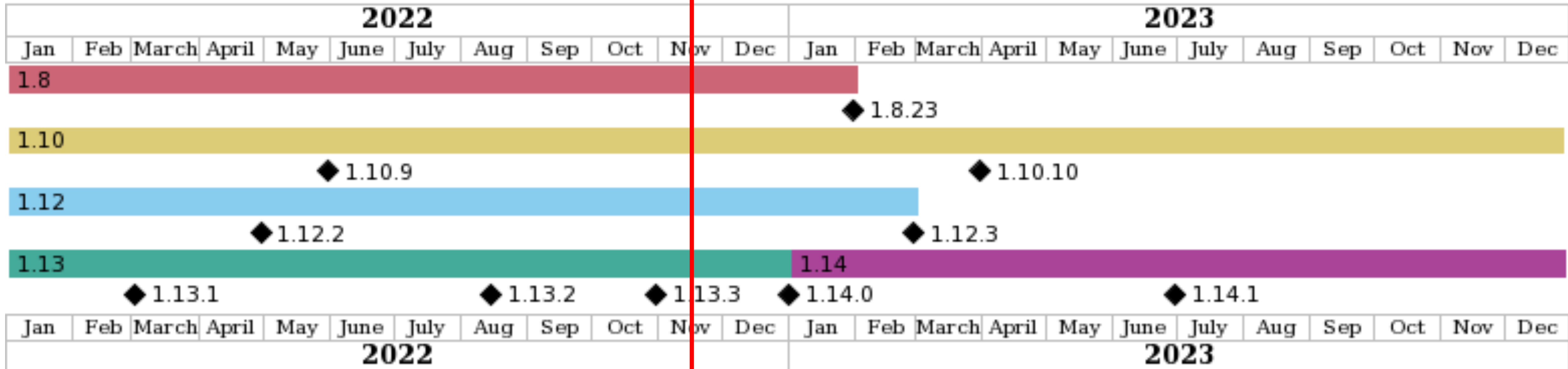


- Release schedule for 2022 - 2023
- HDF5 1.14.0 & new features
- HDF5 2.0.0

Current Release Schedule



HDF5 Release Schedule



<https://github.com/HDFGroup/hdf5>

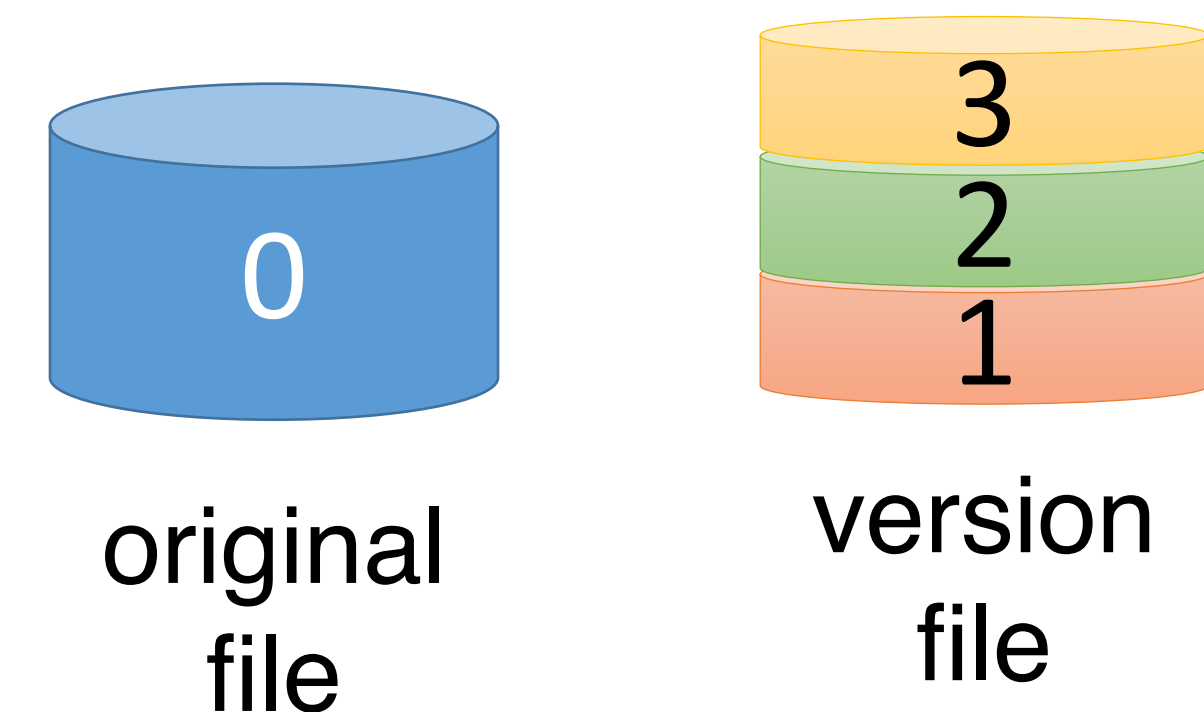
HDF5 1.14.0



- Releases in December 2022
- No new major features since 1.13.3 (released on October 31)
- Library work until then:
 - VOL issue cleanup
 - Addressing some remaining CVEs
 - Documentation
- Includes performance improvements – should be pretty close to 1.8, depending on usage
- Doxygen user guide

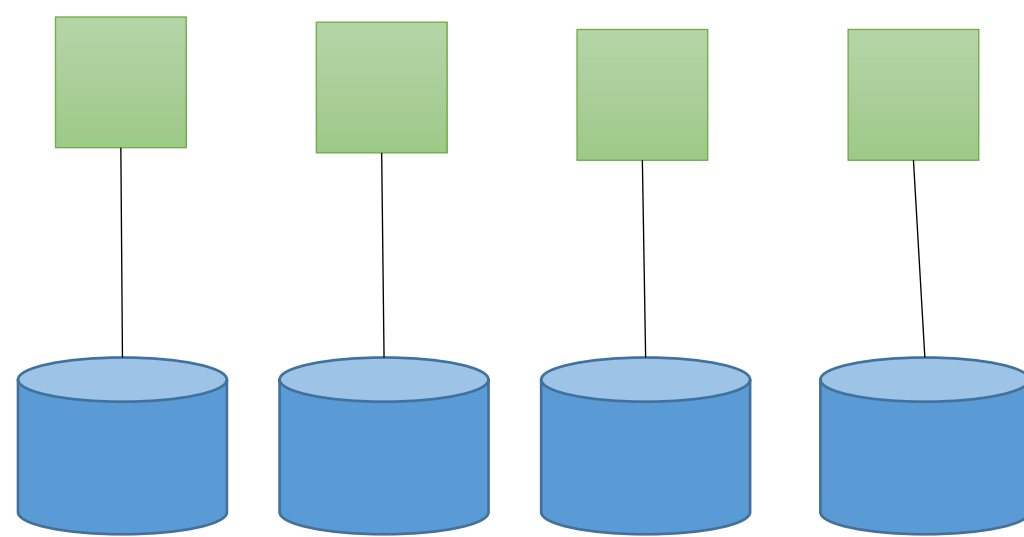
Onion VFD

- Allows for versioned HDF5 files
- Every open/write/close cycle creates a new version
- Original file not touched, but all subsequent versions go in an external "onion" file that contains the new data as well as indexes
- Can open any version for reading, but can only append to the latest version (i.e., no trees)
- Supported by the command-line tools
- Configure with `H5Pset_onion_vfd()`

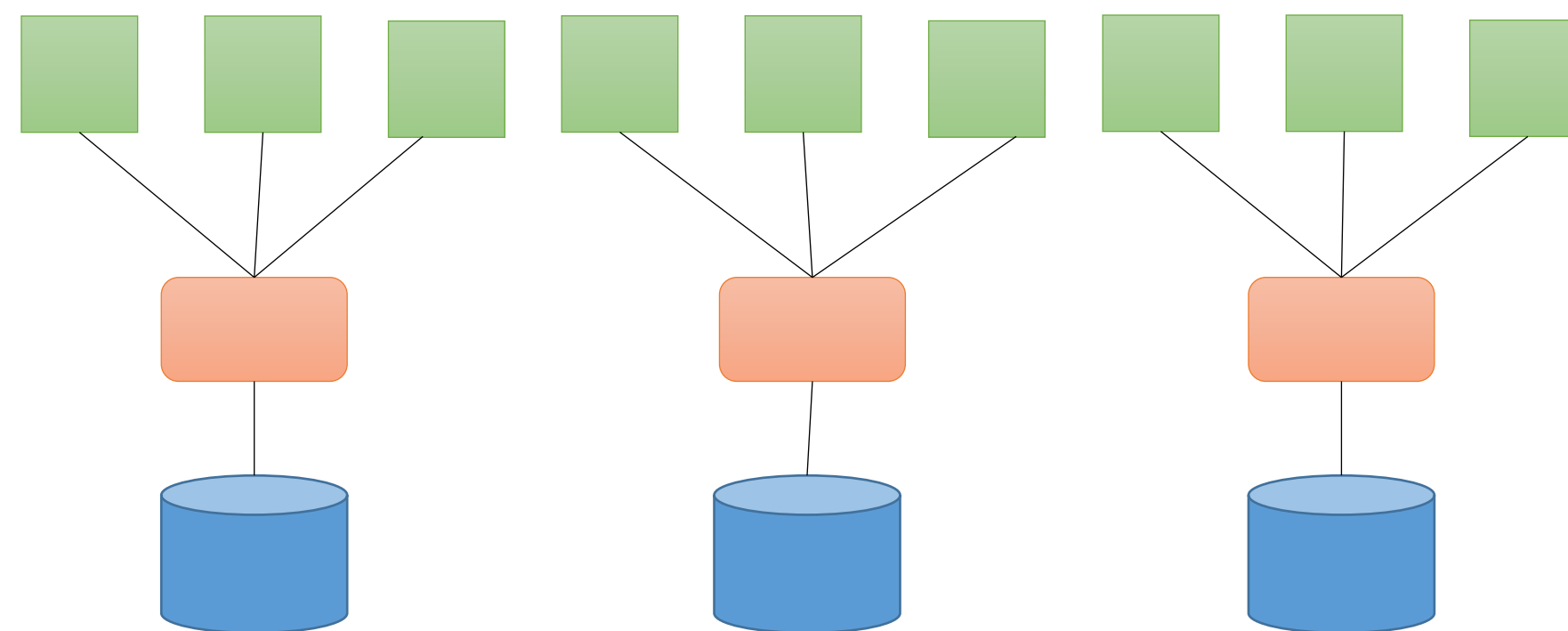


Subfiling

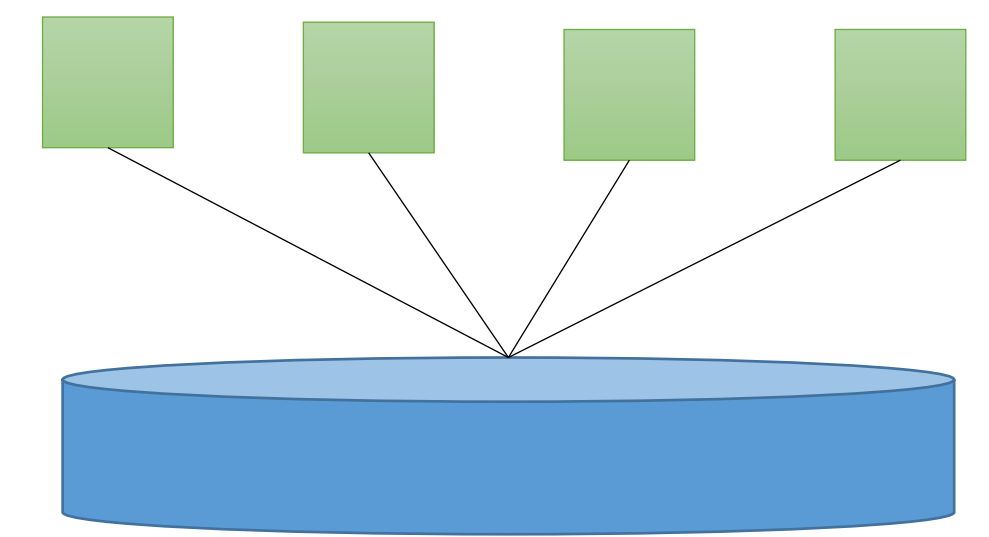
- Implements I/O concentrators
- Find the sweet spot between file-per-process and single-shared-file I/O
- Not built by default, use `--enable-subfiling-vfd` or `HDF5_ENABLE_SUBFILING_VFD`
- Set up with `H5Pset_fapl_subfiling()`



file per process



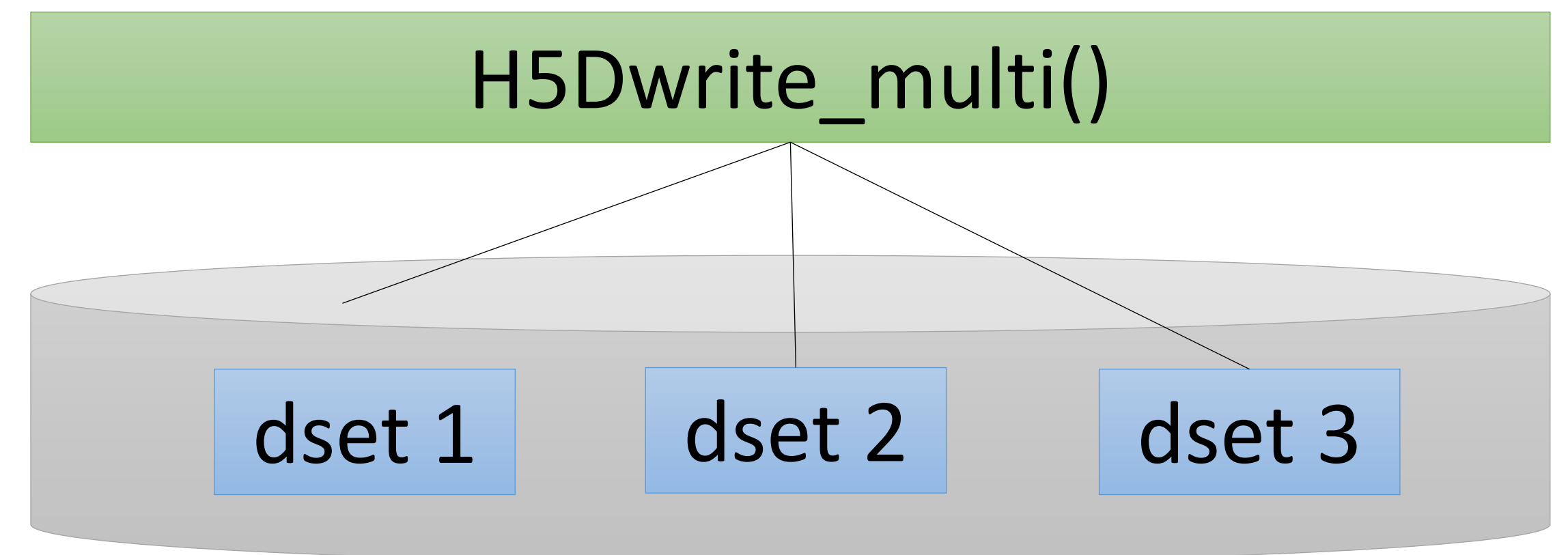
subfiling



single shared file

Multi-dataset I/O

- One dataset I/O call can now modify multiple datasets
- Less API call overhead, allows MPI-I/O et al. to optimize the I/O operation
- Only works on datasets in the same file
- Pass arrays of datasets, dataspace, datatypes, and buffers to `H5Dwrite_multi()` and `H5Dread_multi()` (there are also async versions of these calls)
- Important OpenMPI bug!



Helpful documentation



Onion VFD

https://docs.hdfgroup.org/hdf5/rfc/Onion_VFD_RFC_211122.pdf

<https://www.youtube.com/watch?v=cfshkgr05mA> (demo)

Subfiling

https://docs.hdfgroup.org/hdf5/rfc/RFC_VFD_subfiling_200424.pdf

Selection I/O

https://docs.hdfgroup.org/hdf5/rfc/selection_io_RFC_210610.pdf

Multidataset I/O

https://docs.hdfgroup.org/hdf5/rfc/H5HPC_MultiDset_RW_IO_RFC.pdf

<https://www.hdfgroup.org/2022/11/workarounds-for-openmpi-bug-exposed-by-make-check-in-hdf5-1-13-3/> (bug information)

The Doxygen user guide will be updated before the 1.14.0 release

Updated VOL Interface

- H5VL_class_t struct switches from v2 in 1.13.2 to v3 in 1.13.3 / 1.14.0
- Expanded VOL capability flags
- Vectorized dataset I/O callbacks (to support multi-dataset I/O)

Feature flags RFC

https://github.com/HDFGroup/hdf5doc/blob/master/RFCs/HDF5/VOL/2019-09-23-RFC_VOL_feature_flags.docx

We are in the process of updating the VOL connector author guide and creating a VOL V2 to V3 migration guide.

HDF5 2.0.0



Still working out what will go into this

- Semantic versioning (hence 2.0.0)?
- API "reset" ?
 - Remove deprecated API calls?
 - All calls return an herr_t or hid_t value?
 - No calls return library-allocated memory?
 - ...
- Reworked error interface?
- Drop the Autotools and only support CMake?
- hdim_t type for dataset dimensions?
- Better variable-length storage?
- More pre-defined datatypes?
 - Complex numbers?
 - ML/DL types like float16?
- Reference improvements?
- Change library defaults?
- Drop the C++ wrappers?
- Better Unicode support?
- Many other possibilities!

HDF5 2.0.0



- Looking to have a list of tentative features and a project plan by early spring 2023
- Need to determine a release cadence that balances getting new features out w/ user upgrade pain

Version	Release Date	Gap (Months)
1.0.0	Oct 1998	-
1.2.0	Jul 1999	9
1.4.0	Feb 2001	19
1.6.0	Jul 2003	29
1.8.0	Feb 2008	55
1.10.0	Mar 2016	97
1.12.0	Mar 2020	48
1.14.0	Dec 2023	45
2.0.0	?	?



HDF5 2.0.0



HDF5 is a *community* project, so join the discussion and help us decide!

Most discussions take place on the HDF Group forums:

<https://forum.hdfgroup.org/>

<https://forum.hdfgroup.org/t/what-do-you-want-to-see-in-hdf5-2-0>

<https://forum.hdfgroup.org/t/can-we-retire-the-autotools>

We'll also be creating blog posts, reaching out to our collaborators, hosting town halls (like the Call the Doctor series on Tuesdays), and showing up at conferences like SC for discussions.

<https://zoom.us/meeting/register/tJwvf--gpjsqEtV0NSexRspn0NUjcNhZFmFb>

THANK YOU!

Questions & Comments?